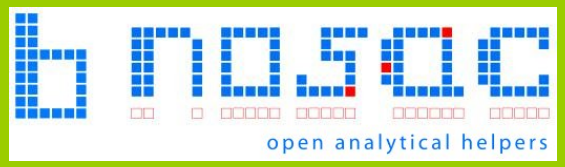
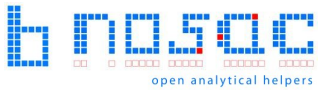
Prediction and Fuzzy Logic at ThomasCook to automate price settings of last minute offers



Jan Wijffels:
jwijffels@bnosac.be

BNOSAC - Belgium Network of Open Source Analytical Consultants
www.bnosac.be

July 10, 2009



Introduction to BNOSAC

- Group of consultants focussed on open source analytical engineering
- Poor man's BI: Python/PostgreSQL/Pentaho/OpenOffice/R...

...

- Expertise in predictive data mining, biostatistics, geostats, python programming, GUI building, artificial intelligence

Business of ThomasCook Belgium

- Sell holidays (sun and beach in this user case)
- 70 destinations around Mediterranean and Americas

- Own planes & bought seats need to be filled with passengers
- Flight frequence for some destinations up to 4 flights within one day. Some flights can be combined (BRU->ACE->FUE->ACE->BRU)

Introduction to last minute price settings

- Last minute prices departures Brussels/Liège/Ostend/Lille
- Up to 2 months before departure
- People book now to go on holiday e.g. August 10, 2009 to destination X. Can stay 3-28 nights, choose among several hotels, with certain board (All Inclusive, B&B, ...) and certain room type.

e.g. Hurghada (HRG): dayly flights from Brussels (BRU)

\# prices in August: 31 days $\times$ 12 durations $\times$ 2 brands $\times$ 20 hotels $\times$ 4 boards $\times$ 3 room types $= \pm 248000$ prices

- Prices can go $\nearrow$ or $\searrow$ depending on offer and demand

Business challenge

Business challenge

Fill the planes at the highest prices so that the plane doesn't fill too fast and make sure all seats are filled.

- Currently 2.9 Mio promotional prices on the market. Prices change dayly.
- Only cover approaches towards prices of packages (flight + hotel), only price effects of couples (so no children).

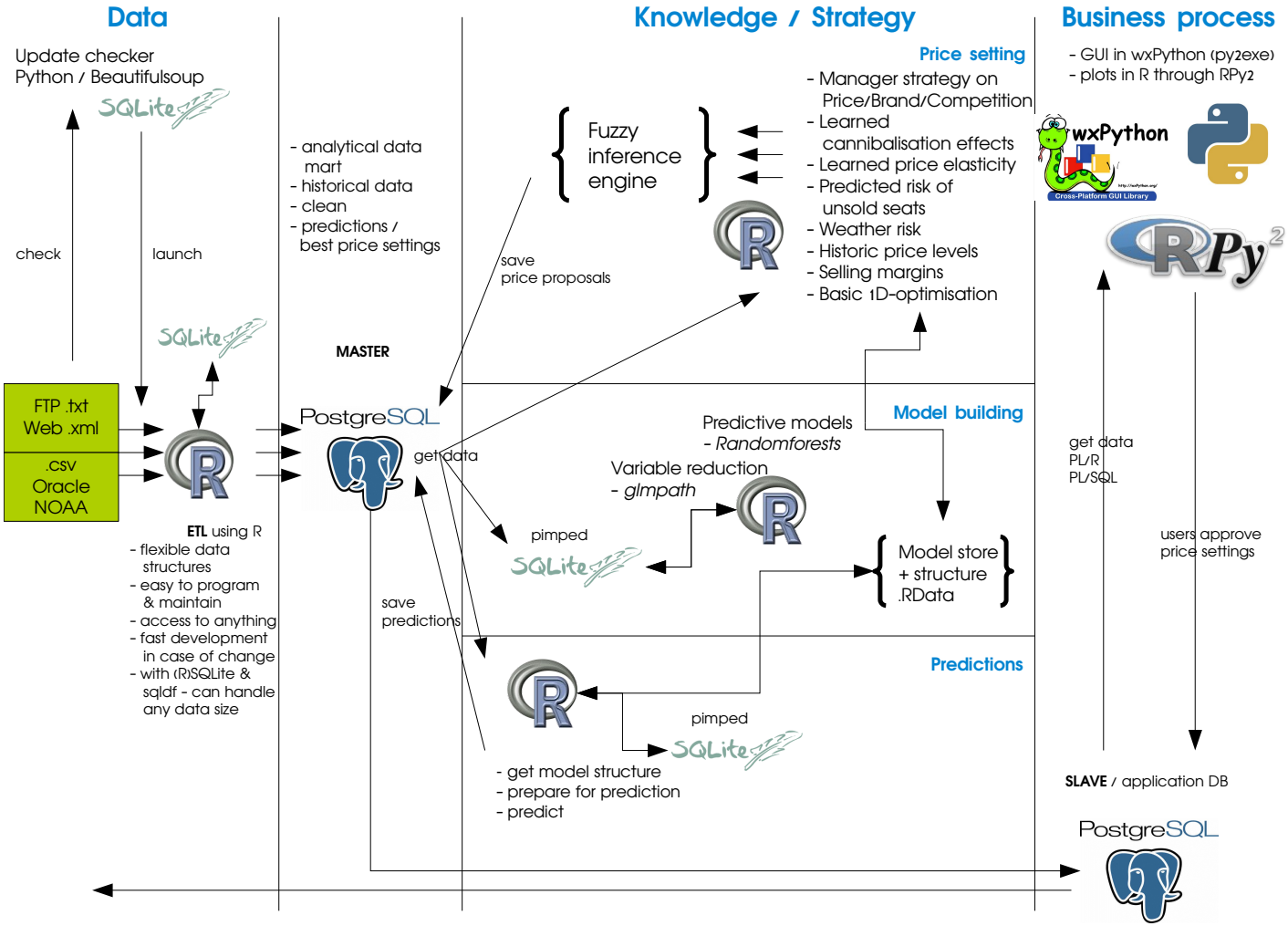
Optimisation problem

- A lot of factors influencing bookings:
  - Holiday information / Day of the week
  - Flight information (hours of departure and of return flights, availability of flights)
  - Weather
  - Prices (2 brands, competitor) and price evolution
  - Cannibalisation (risk of losing passengers to yourself)
    - prices of similar destinations - last minute customers only want the sun at the cheapest price
    - prices on similar departure dates (a few days later/earlier)
  - Days before departure
  - ... dimensionality is large ($>$ 100000 factors could influence bookings on flight from BRU to HRG on August 10, 2009)
- Find the best price settings over all these parameters to ...
  - optimize margin / minimize risk / optimize market share

Data & speed challenge

-  Data size last year only
  - own last minute promotional prices: >450 million records.
  - competitor prices
  - flight info: ± 60000 flights on the market × 365 days ± 21.900.000 records
  - weather info at noon:
    70 destinations × 365 days × weather forecasts
- Speed
  -  "Hello prices" at ±7o'clock in the morning (mainframe).
     "Hello employees" at ±8h30 in the morning
  - ±1h30 to make predictions and give 'best' automatic price proposals

Architectural solution



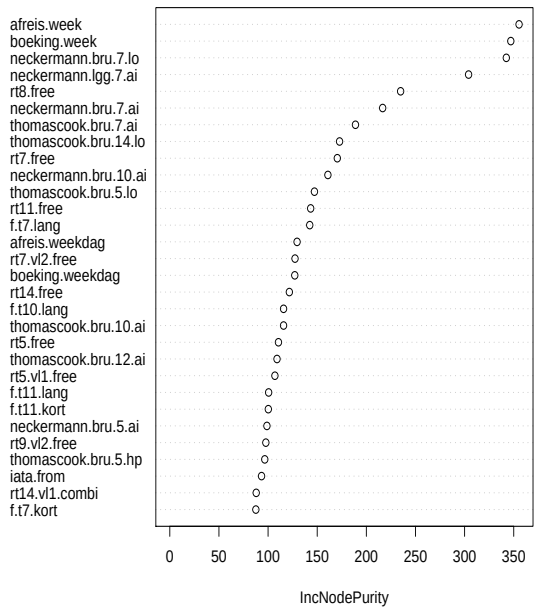
Analytical solution: Predictive modelling

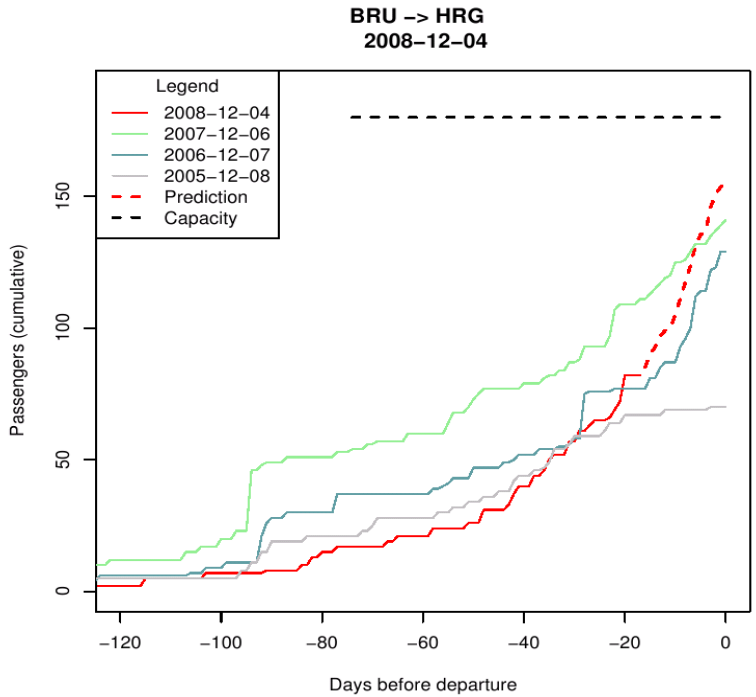
Out of the box solutions exist in R. 'Best practice' approach:

- Pimp SQLite so that it can handle tables with up to ±30000 columns. Raw model tables dim 20.000.000 x 30000
- Data preparation (missing values, split numeric data in categories) - do heavy reshaping/juggling/merging/indexing in (R)SQLite & sqldf, use R for advanced data features
- Sample depending on CPU/RAM and statistical technique: we have 4 dual cores, 64bit Linux, 32Gb RAM.
- Reduce: GLM with penalization on the size of the L1 norm of the coefficients $L(\beta, \lambda) = -\sum_{i=0}^{n} y_i \theta(\beta)_i b(\theta(\beta)_i) + \lambda \|\beta\|_1$ (glmpath package)

Analytical solution: Predictive modelling cont.

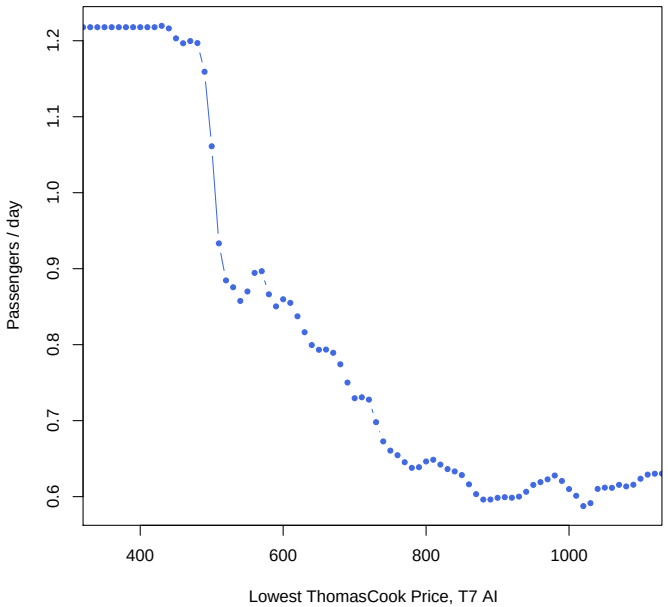
- Only most important predictors to build randomForest
- Use randomForest model to predict how fast the flights will fill.

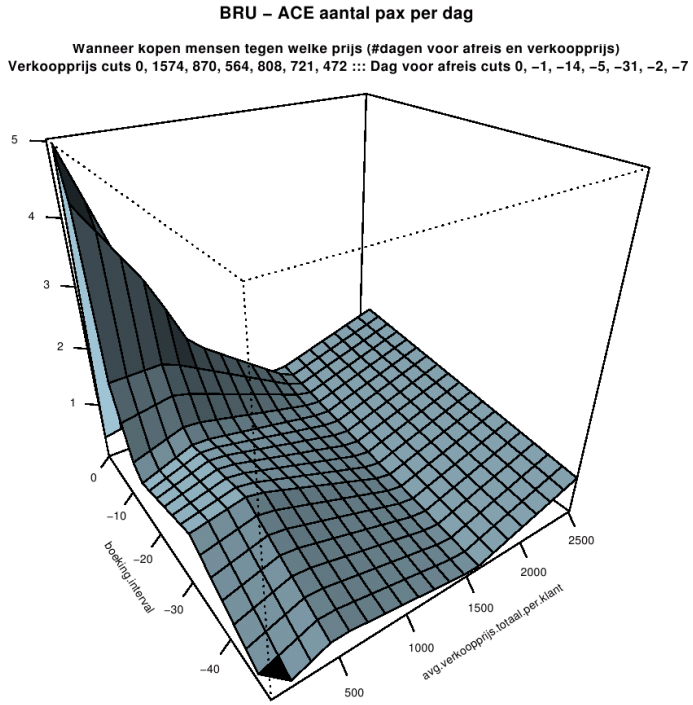




Analytical solution: Predictive modelling cont.

- Get the price effects from the randomForest model and use it:
- Do fast 1- or 2-dimensional optimisation to fill seats that will not be filled according to the forecast at the optimal price.



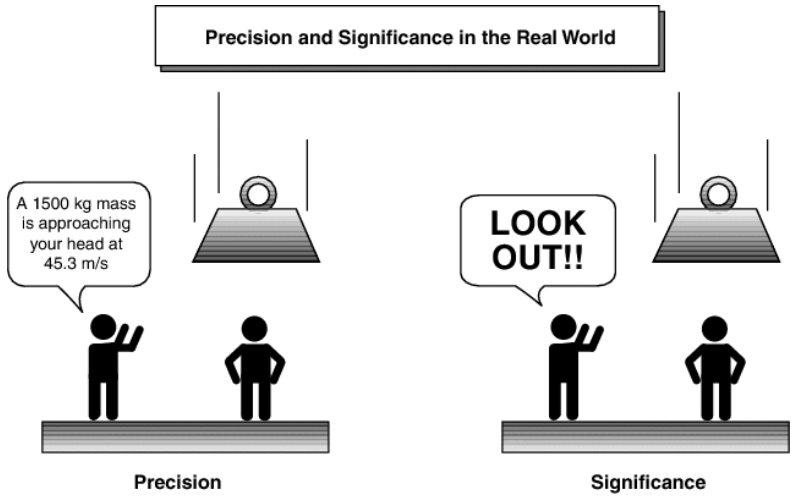


Analytical solution: Fuzzy Logic

Prediction and optimisation is nice but not enough

Managers reason with words/concepts. Mimic them and combine their logic with predictive logic. How?

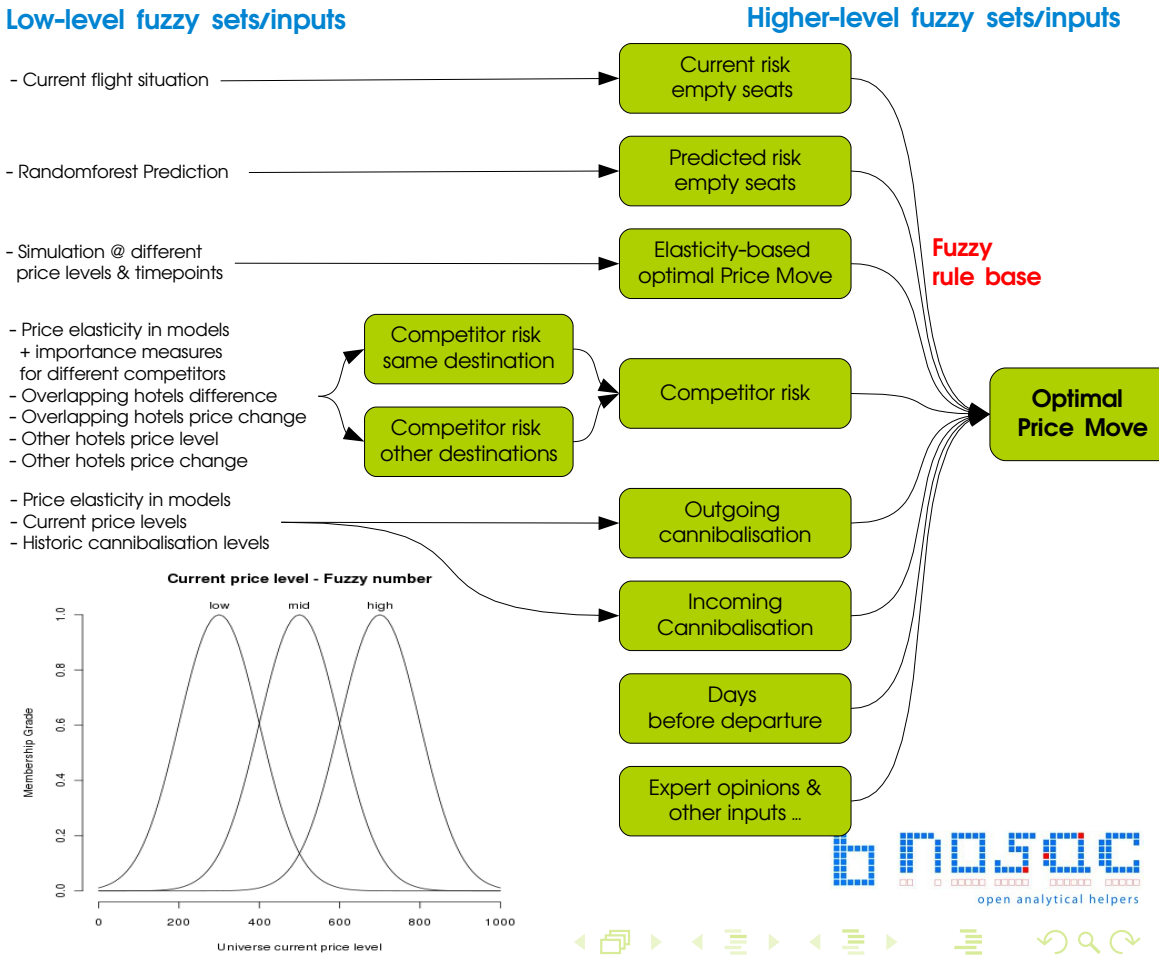
- Map business concepts to fuzzy sets.
- Make fuzzy rule-based engine reflecting how managers/business users decide on price settings
- Do fuzzy inference to obtain new price settings



Analytical solution: Fuzzy Logic cont.

Map business concepts to fuzzy sets.

- Listen to the people. Fuzzy concepts have blurred boundaries.
- Map linguistic variables to a membership degree $\mu(x) \in [0, 1]$
- sets package (Hornik K., Meyer D., Buchta C.)
- fuzzy_normal, fuzzy_trapezoid, fuzzy_sigmoid, ...



Analytical solution: Fuzzy Logic cont.

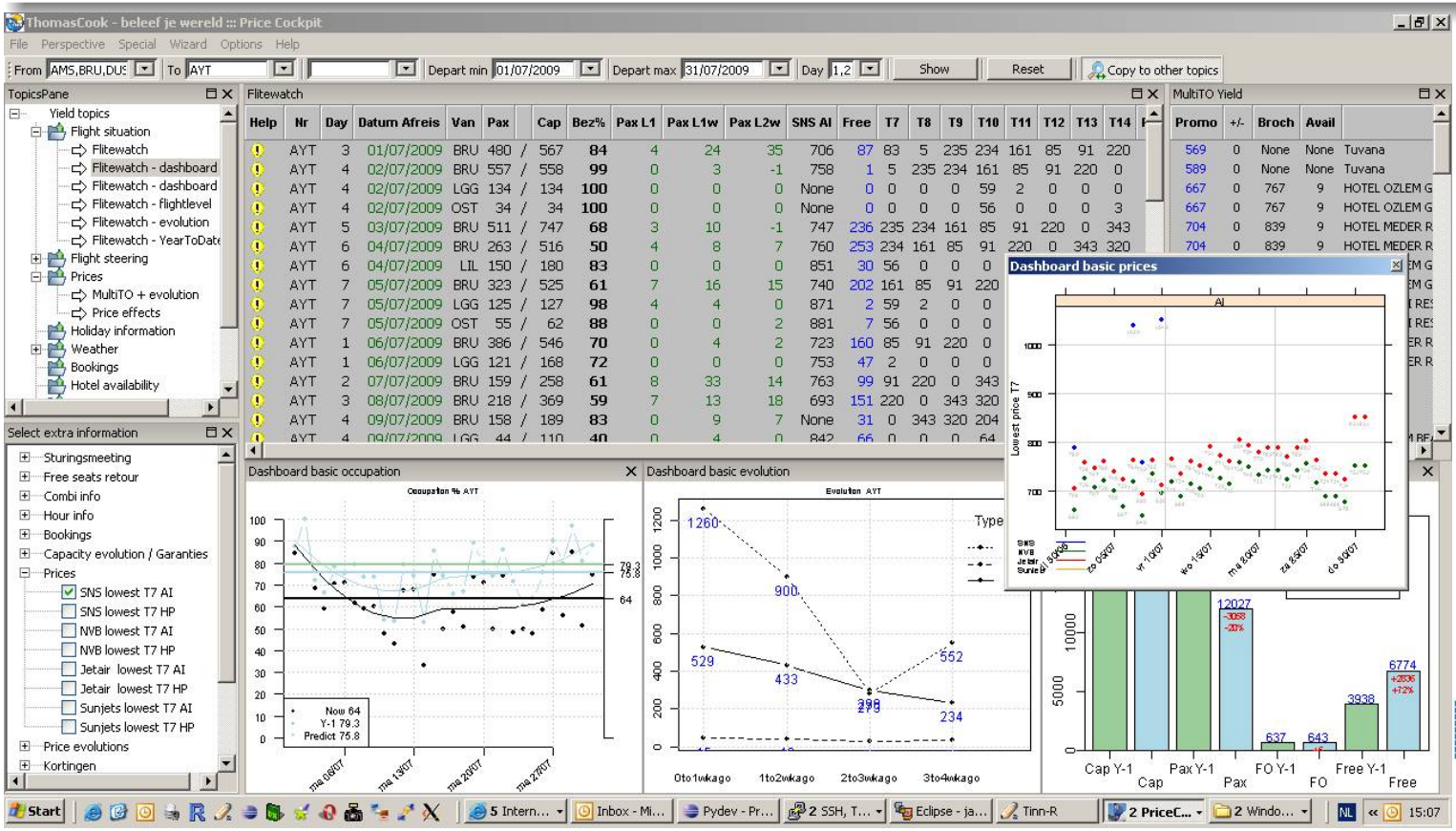
Make fuzzy rule-based engine, do fuzzy inference & defuzzify.

```
rules <- set(
fuzzy_rule(predicted_risk %is% low, price_change %is% up),
fuzzy_rule(predicted_risk %is% high
  & competitor_risk %is% high, price_change %is% down_high)
...)
simple.system <- fuzzy_system(variables, rules)
fuzzy.best.price <- fuzzy_inference(simple.system, NEWDATA)
gset_defuzzify(fuzzy.best.price, "centroid")
```

- Different business strategies can be easily mapped to fuzzy inference engines.

Influence the business process, use visuals, build GUI

Prediction, optimisation and improving on business users
is nice but not enough, you need to influence the business process



PL/R, RPy2, GUI's in R, people

- PL/R.
  - Had a lot of shared memory problems while other processes were runnning. But probably overkilled it (run PL/R script which calls some R code from within R process that uses RdbiPgSQL)
  - Debugging hell.
  - R & SQLite is our best choice for heavy data juggling.
  - PL/R is OK for collecting information on diverse data sources in 1 call from a remote machine.
  - Useful for plotting purposes in SaaS framework.

PL/R, RPy2, GUI's in R, people cont.

- User interfaces - developer view
  - Combining wxPython and R through RPy2 is easy and simple.
  - py2exe gives easy python binary executables, people only need to have R installed to access its power
- User interfaces - IT view
  - IT departments don't like R
  - R should be SaaS, central server where people can connect to
- User interfaces - business user point of view
  - They don't care about R
  - GUI and plotting the results helped convincing them
  - Fuzzy logic allowed them to interact and stick to the business.
  - Combining the results with an improved business process was the most convincing factor.

Questions?

http://www.bnosac.be